

A simulation-based approach for solving optimisation problems with ODE-type steady state constraints ^{*}

Anna Fiedler[†]

Fabian J. Theis[†]

Jan Hasenauer[†]

Abstract

Ordinary differential equations (ODEs) are widely used to model biological, (bio-)chemical and technical processes. The parameters of these ODEs are often estimated from experimental data using ODE-constrained optimisation. This article proposes a simple simulation-based approach for solving optimisation problems with steady state constraints relying on an ODE. This simulation-based optimisation method is tailored to the problem structure and exploits the local geometry of the steady state manifold and its stability properties. A parameterisation of the steady state manifold is not required. We prove local convergence of the method for locally strictly convex objective functions. Efficiency and reliability of the proposed method are demonstrated in two examples. The proposed method demonstrated better convergence properties than existing general purpose methods and a significantly higher number of converged starts per time.

1 Introduction

Ordinary differential equation models are widely used in engineering, physics and life sciences. As the parameters of the underlying processes are often unknown, they have to be determined from experimental data [1]. This requires efficient and robust parameter optimisation methods. Unfortunately, for a wide range of models these efficiency and robustness requirements are not met by available methods (see discussion [2]). In particular parameter estimation and optimisation problems with nonlinear equality constraints, e.g., steady state constraints, are intricate.

In the following we will consider ODE models

$$\frac{dx}{dt} = f(\theta, x), \quad x(0) = x_0 \quad (1)$$

with state variables $x(t) \in \mathbb{R}^{n_x}$ and parameters $\theta \in \mathbb{R}^{n_\theta}$. The vector field $f : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$ is assumed to be Lipschitz continuous to ensure existence and uniqueness of the solution. The parameter dependent steady states $x_s(\theta)$ of such systems are constrained by

$$0 = f(\theta, x_s). \quad (2)$$

This defines a nonlinear manifold. As many systems operate in steady state and/or undergo fast equilibration, constrained optimisation problems

$$\begin{aligned} \min_{x_s, \theta} J(\theta, x_s) \\ \text{s.t. } f(\theta, x_s) = 0, \end{aligned} \quad (3)$$

with objective function $J : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_x} \rightarrow \mathbb{R}$, are widely studied. Such constrained optimisation problems arise, e.g., in systems biology when model parameters are estimated from steady state data and in engineering when the steady state of a process is designed.

A variety of methods has been developed for optimisation under nonlinear equality constraints (3). In particular local and manifold optimisation are used in practice [3, 4]. Standard local optimisers have to move on the nonlinear

^{*}The authors acknowledge financial support from the Postdoctoral Fellowship Program (PFP) of the Helmholtz Zentrum München and the German Federal Ministry of Education and Research (BMBF, Grant no. 01ZX1310B) within the SYS-Stomach project.

[†]Institute of Computational Biology, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany; and Chair of Mathematical Modeling of Biological Systems, Center for Mathematics, Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany. anna.fiedler@tum.de, fabian.theis@helmholtz-muenchen.de, jan.hasenauer@helmholtz-muenchen.de

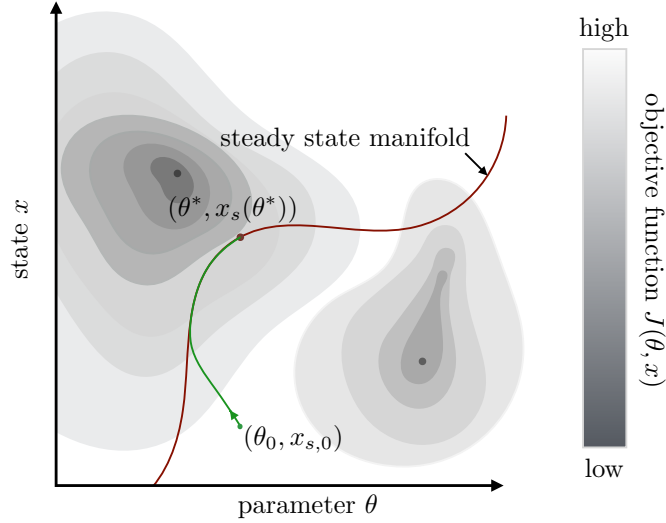


Figure 1: Schematic illustration of steady state constrained optimisation problem. The path of a continuous analogue of a constrained optimiser (—) from a starting point $(\theta_0, x_{s,0})$ to a local minimum $(\theta^*, x_s(\theta^*))$ on the steady state manifold (—) is depicted.

manifold defined by the steady state constraint (Figure 1). These methods become quickly inefficient and/or have convergence problems [4]. Improvements are achieved by reformulating the constrained to an unconstrained optimisation problem using Lagrangian multipliers. However, this increases the problem size. Algorithms for manifold optimisation generally circumvent an increase of the problem size and moves on the manifold. Instead retraction operators are employed [3], projecting a point onto the manifold. Retraction operators which can be efficiently evaluated are unfortunately only available for a few classes of manifolds. For vector fields f which are non-linear, there are generally no analytical functions enabling the projections, limiting the application of established algorithms for optimisation on manifolds.

In the following, we will propose a simulation-based framework for numerical optimisation with steady state constraints. Using ideas from continuous analogues of optimisers [5, 6, 7] and systems theory [8], we will introduce an ODE system exploiting the local geometry of the steady state manifold to move on the manifold. The manifold will be stabilised using a continuous retraction, available because we are interested in stable steady states of the model (1). The simulation of the resulting ODE system with sophisticated adaptive step-size solvers yields an improved optimisation performance. We provide a proof for exponential stability of a locally optimal point for a locally strictly convex objective function $J(\theta, x)$ and a locally exponentially stable system (1). To evaluate the developed optimisation method, two application examples from the field of systems biology are studied.

2 Geometry of the steady state manifold

To develop a tailored method for solving optimisation problems with steady state constraints, we will exploit the first order geometry of the manifold of steady states. To this end we consider the sensitivities of the states x with respect to the parameters θ ,

$$s_i := \frac{\partial x}{\partial \theta_i} = \left(\frac{\partial x_1}{\partial \theta_i}, \dots, \frac{\partial x_{n_x}}{\partial \theta_i} \right)^T,$$

whose dynamics are governed by the forward sensitivity equation

$$\dot{s}_i = \nabla_x f s_i + \nabla_{\theta_i} f, \quad s_i(0) = \nabla_{\theta_i} x_0.$$

In matrix form, we obtain

$$\dot{S} = \nabla_x f S + \nabla_{\theta} f, \quad S(0) = \nabla_{\theta} x_0, \quad (4)$$

with $S(t) = (s_1(t), s_2(t), \dots, s_{n_{\theta}}(t)) \in \mathbb{R}^{n_x \times n_{\theta}}$.

In a steady state $x_s(\theta)$ corresponding to some parameter θ , the forward sensitivity equation (4) simplifies to

$$0 = \nabla_x f S + \nabla_\theta f \quad (5)$$

evaluated at $(\theta, x_s(\theta))$. If the considered steady state $x_s(\theta)$ is locally exponentially stable, the Jacobian $\nabla_x f|_{(\theta, x_s(\theta))}$ is negative definite and invertible. Accordingly,

$$S = -(\nabla_x f)^{-1} \nabla_\theta f. \quad (6)$$

Furthermore, this implies that there is a local parametrization of the steady state, $x_s(\theta)$ such that $(\theta, x_s(\theta))$ is an isolated root of (2). The sensitivity of the steady state with respect to the parameters, S , can be used in the Taylor series of $x_s(\theta)$,

$$x_s(\theta + r\Delta\theta) = x_s(\theta) + Sr\Delta\theta + O(r^2), \quad (7)$$

with $S = S(\theta, x_s(\theta))$. The perturbation direction and the step size are denoted by $\Delta\theta$ and r , respectively. The Taylor expansion (7) reveals how the steady state changes locally with the parameters. By reformulating (7) and letting $r \rightarrow 0$, we obtain a dynamical system which evolves on $x_s(\theta)$,

$$\frac{dx_s}{dr} = S\Delta\theta. \quad (8)$$

Given an update direction $\Delta\theta$ and a length r , (8) provides the steady state $x_s(\theta + r\Delta\theta)$ up to the accuracy of the used ODE solver. Hence, (8) enables moves on the steady state manifold, similar to results in [3]. A recalculation of the steady state for changed parameter values or a retraction operator are not longer required.

3 Optimisation with steady state constraints

In this section, we exploit the insight into the local geometry of the steady state manifold to derive an ODE model converging to a locally optimal solution of the constrained optimisation problem (3).

3.1 Simulation-based optimisation method.

In local optimisation, the parameters θ are updated using the gradient of the objective function. The method of gradient descent, for instance, exploits an update

$$\theta_{k+1} = \theta_k - r \frac{dJ}{d\theta}^T, \quad (9)$$

with step length r . The continuous analogue of gradient decent methods is $d\theta/dr = -(dJ/d\theta)^T$ [7]. This ODE system can be coupled with the dynamical system evolving on the steady state manifold (8), using $\Delta\theta = -(dJ/d\theta)^T$. We obtain the ODE system

$$\begin{aligned} \frac{d\theta}{dr} &= - \left. \frac{dJ}{d\theta} \right|_{(\theta, x_s)}^T \\ \frac{dx_s}{dr} &= S \frac{d\theta}{dr} = -S \left. \frac{dJ}{d\theta} \right|_{(\theta, x_s)}^T, \end{aligned} \quad (10)$$

with the steady state sensitivity S . Initialisation of (10) in a point $(\theta_0, x_{s,0})$ fulfilling (2) yields a trajectory evolving on the steady state manifold, along which the objective function decreases.

The formulation (10) bears two disadvantages: (i) An appropriate initial point $(\theta_0, x_{s,0})$ has to be determined by solving (2); and (ii) numerical errors can result in a divergence from the steady state manifold. To address these problems, we introduce the term $\lambda f(\theta, x_s)$ which locally retracts the state of the system to the manifold by exploiting the stability properties of the steady state. This yields the system

$$\begin{aligned} \frac{d\theta}{dr} &= - \left. \frac{dJ}{d\theta} \right|_{(\theta, x_s)}^T \\ \frac{dx_s}{dr} &= -\hat{S} \left. \frac{dJ}{d\theta} \right|_{(\theta, x_s)}^T + \lambda f(\theta, x_s). \end{aligned} \quad (11)$$

For this modified system we do not require that the initial point $(\theta_0, x_{s,0})$ fulfils the steady state conditions (2), hence, the Jacobian $\nabla_x f|_{(\theta, x_s)}$ might not be invertible. To address this, we define

$$\hat{S}(\theta, x_s) := -(\nabla_x f|_{(\theta, x_s)})^+ \nabla_\theta f|_{(\theta, x_s)}, \quad (12)$$

in which $(\nabla_x f|_{(\theta, x_s)})^+$ denotes the Moore–Penrose pseudoinverse of $\nabla_x f|_{(\theta, x_s)}$. On the steady state manifold, the Jacobian is invertible and we recover the standard steady state sensitivity. For large retraction factor $\lambda \gg 0$, the state (θ, x_s) is retracted quickly to the steady state manifold $(\theta, x_s(\theta))$.

3.2 Analysis of simulation-based optimisation method.

In the following, we assess the local stability and convergence of (11) to a local optimum $(\theta^*, x_s(\theta^*))$ under four conditions.

Assumption 3.1. *There exists an locally isolated root $x_s(\theta)$ of the equation $f(\theta, x_s(\theta)) = 0$ for all $(\theta - \theta^*) \in B_c = \{\tilde{\theta} \in \mathbb{R}^{n_\theta} \mid \|\tilde{\theta}\| \leq c\} \subseteq \mathbb{R}^{n_\theta}$.*

This root is a steady state of the system (1).

Assumption 3.2. *The steady state $x_s(\theta)$ is locally exponentially stable uniformly in θ , with $(\theta - \theta^*) \in B_c$.*

Accordingly, $\exists \zeta_0, \gamma, k > 0$ such that $\|x(t) - x_s(\theta)\| \leq k\|x(0) - x_s(\theta)\| \exp(-\gamma t)$, $\forall (x(0) - x_s(\theta)) \in B_{\zeta_0} := \{\tilde{\Delta}_s(0) \in \mathbb{R}^{n_x} \mid \|\tilde{\Delta}_s(0)\| < \zeta_0\}$ and $\forall (\theta - \theta^*) \in B_c$. This implies that $\exists \zeta \geq \zeta_0$ such that $(x(t) - x_s(\theta)) \in B_\zeta := \{\tilde{\Delta}_s(t) \in \mathbb{R}^{n_x} \mid \|\tilde{\Delta}_s(t)\| < \zeta\}$ [8].

Assumption 3.3. *There exists a neighbourhood around θ^* , $(\theta - \theta^*) \in B_c$, in which the objective function evaluated on the steady state manifold $x_s(\theta)$, $J(\theta, x_s(\theta))$, is locally strictly convex in θ .*

This implies that the parameters θ are locally structurally identifiable and that solutions to $d\theta/dr = -dJ/d\theta|_{(\theta, x_s(\theta))}$ converge to the optimal point θ^* for initial points $(\theta(0) - \theta^*) \in B_c$.

Assumption 3.4. *The functions $(dJ/d\theta)|_{(\theta, x_s)}$, $f(\theta, x_s)$ and $x_s(\theta)$ and their partial derivatives up to order 2 are bounded for $(\theta - \theta^*) \in B_c$ and $(x_s - x_s(\theta)) \in B_\zeta$.*

Using these assumptions we find:

Theorem 3.1. *Let Assumptions 3.1 - 3.4 be satisfied. Then there exists a λ^* such that for all $\lambda > \lambda^*$ a local minimum (θ^*, x_s^*) of the optimisation problem (3) is a locally exponentially stable steady state of the system (11).*

Proof. To prove Theorem 3.1 we use perturbation theory [8]. We define $\varepsilon = \lambda^{-1}$ and shift the optimum to the origin using the linear state transformation,

$$\tilde{\theta} = \theta - \theta^* \quad \text{and} \quad \tilde{x}_s = x_s - x_s(\theta^*).$$

This yields the singular perturbed system

$$\begin{aligned} \frac{d\tilde{\theta}}{dr} &= -\frac{dJ^T}{d\tilde{\theta}} =: \tilde{F}(\tilde{\theta}, \tilde{x}_s, \varepsilon) \\ \varepsilon \frac{d\tilde{x}_s}{dr} &= -\varepsilon \hat{S} \frac{dJ^T}{d\tilde{\theta}} + f =: \tilde{G}(\tilde{\theta}, \tilde{x}_s, \varepsilon), \end{aligned} \quad (13)$$

with $dJ/d\tilde{\theta}$, f and \hat{S} evaluated at $(\tilde{\theta} + \theta^*, \tilde{x}_s + x_s(\theta^*))$. The system (13) captures the dynamics of the deviances from the optimal parameter, $\tilde{\theta}$, and the deviances from the steady state for the optimal parameter, \tilde{x}_s . Furthermore, it possesses the following properties:

- (i) $\tilde{F}(0, 0, \varepsilon) = 0$ and $\tilde{G}(0, 0, \varepsilon) = 0$ as the objective function gradient vanishes, $0 = (dJ/d\tilde{\theta})|_{(\theta^*, x_s(\theta^*))}^T$ and as the steady state condition is fulfilled, $0 = f(\theta^*, x_s(\theta^*))$. Both follows from optimality of θ^* and Assumption 3.1.

- (ii) The equation $0 = \tilde{G}(\tilde{\theta}, \tilde{x}_s, 0)$ has the isolated root $\tilde{x}_s(\tilde{\theta}) = x_s(\tilde{\theta} + \theta^*) - x_s(\theta^*)$ (Assumption 3.1) with $\tilde{x}_s(0) = 0$.
- (iii) The functions \tilde{F} , \tilde{G} , and x_s and their partial derivatives up to order 2 are bounded for $(\tilde{x}_s - x_s(\tilde{\theta})) \in B_\zeta$ (Assumption 3.4).
- (iv) The origin of the reduced systems

$$\frac{d\tilde{\theta}}{dr} = F(\tilde{\theta}, \tilde{x}_s(\tilde{\theta}), \varepsilon) = - \left. \frac{dJ}{d\tilde{\theta}} \right|_{(\tilde{\theta} + \theta^*, x_s(\tilde{\theta}) + x_s(\theta^*))}^T$$

obtained for $\varepsilon = 0$ is locally exponentially stable, with $x_s(\tilde{\theta}) + x_s(\theta^*) = x_s(\tilde{\theta} + \theta^*)$ from (ii). This follows from strict local convexity of the objective at θ^* (Assumption 3.3).

- (v) The boundary-layer system is derived from (13) using the transformation $\tilde{\Delta}_s = \tilde{x}_s - \tilde{x}_s(\tilde{\theta}) (= \tilde{x}_s - \tilde{x}_s(\tilde{\theta}))$, yielding

$$\begin{aligned} \frac{d\tilde{\theta}}{dr} &= - \frac{dJ}{d\tilde{\theta}}^T \\ \varepsilon \frac{d\tilde{\Delta}_s}{dr} &= \varepsilon(S - \hat{S}) \frac{dJ}{d\tilde{\theta}}^T + f, \end{aligned}$$

with $dJ/d\tilde{\theta}$, f and \hat{S} evaluated at $(\tilde{\theta} + \theta^*, \tilde{\Delta}_s + x_s(\tilde{\theta} + \theta^*))$ and S evaluated at $(\tilde{\theta} + \theta^*, x_s(\tilde{\theta} + \theta^*))$. After rescaling of the simulation time, $\rho = r/\varepsilon$, and setting $\varepsilon = 0$, we obtain the boundary-layer system

$$\frac{d\tilde{\Delta}_s}{d\rho} = f(\tilde{\theta} + \theta^*, \tilde{\Delta}_s + x_s(\tilde{\theta} + \theta^*)).$$

The origin of this boundary-layer system is locally exponentially stable, uniformly in $\tilde{\theta}$ as the steady state $x_s(\theta)$ of (1) is exponentially stable uniformly in $\theta = \tilde{\theta} + \theta^*$ (Assumption 3.2).

The properties (i)-(v) are the prerequisites of [8, Theorem 9.3], establishing the existence of $\varepsilon^* > 0$ such that for all $\varepsilon < \varepsilon^*$ systems of type (13) are locally exponentially stable. As stability properties are conserved under the performed transformations, we obtain for $\varepsilon = \lambda^{-1}$ the Theorem 3.1. \square

Theorem (3.1) establishes local exponential stability of local optima $(\theta^*, x_s(\theta^*))$ for appropriate choice of λ , assuming a convex objective function (3) and a well-behaved system (1). From this follows also local convergence. Stability and convergence are not affected by the approximation of the steady state sensitivity via \hat{S} .

3.3 Analysis of multiple steady-state constraints.

In practice often steady state data for different inputs u_i are available, yielding constraints $f(\theta, x_s^i, u_i) = 0$ for $i = 1, \dots, m$. The simulation-based method and the stability proof presented in Sections 3.1 and 3.2 can be generalised to this case. We simply concatenate constraints and state vectors, $\tilde{f}(\theta, x) = (f(\theta, x_s^1, u_1), \dots, f(\theta, x_s^m, u_m))$ with $x = (x_s^1, \dots, x_s^m)$, and apply the method to the concatenated system. Due to the block structure some calculations, e.g., matrix inversion, can be simplified.

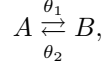
In the following, we illustrate the behaviour of the simulation-based optimisation method. Furthermore, the performance of the proposed approach will be compared to standard constrained and unconstrained optimisation methods. For this purpose, we consider two simulation examples for which the ground truth is known.

4 Example 1: Conversion reaction

In this section, we study parameter estimation for conversion reactions from steady state data. Conversion reactions are among the most common motifs in biological systems, therefore particularly interesting, and provide a simple test case.

4.1 Pathway model.

We consider the conversion reaction



with parameters $\theta = (\theta_1, \theta_2) \in \mathbb{R}_+^2$. It is well-known that the concentration of the biochemical species A is governed by

$$\frac{dx}{dt} = \theta_2 \xi - (\theta_1 + \theta_2)x, \quad x(0) = x_0 \quad (14)$$

with $x_0 \in \mathbb{R}_+$ denoting the initial concentration of A and $\xi \in \mathbb{R}_+$ denoting the sum of the initial concentrations of A and B . The steady state of model (14) is

$$x_s(\theta) = \frac{\theta_2 \xi}{\theta_1 + \theta_2}. \quad (15)$$

4.2 Parameter estimation problem.

To illustrate the properties of the simulation-based method, we consider a simple parameter estimation problem. We assume that for $\xi = 1$, the steady state $\bar{x}_s = 0.2$ is observed. Furthermore, for the parameters θ_i , $i = 1, 2$, prior knowledge suggests the value $\bar{\theta}_1 = 3.9$ and $\bar{\theta}_2 = 1.5$. Accordingly, we estimate the parameters using the weighted least squares optimisation problem

$$\begin{aligned} \min_{\theta, x_s} J(\theta, x_s) &:= \frac{1}{2} 10 (x_s - \bar{x}_s)^2 + \frac{1}{2} \sum_{i=1}^2 (\theta_i - \bar{\theta}_i)^2 \\ \text{s.t. } 0 &= \theta_2 \xi - (\theta_1 + \theta_2)x_s. \end{aligned} \quad (16)$$

4.3 Formulation of system for simulation-based optimisation.

To formulate the ODE model solving (16), the gradient of the objective function with respect to the parameters

$$\begin{aligned} \frac{dJ}{d\theta_1} &= 10(x_s - \bar{x}_s)s_1 + (\theta_1 - \bar{\theta}_1) \\ \frac{dJ}{d\theta_2} &= 10(x_s - \bar{x}_s)s_2 + (\theta_2 - \bar{\theta}_2) \end{aligned}$$

and the local sensitivities

$$S(\theta, x_s) = (s_1, s_2) = \left(\frac{-x_s}{\theta_1 + \theta_2}, \frac{\xi - x_s}{\theta_1 + \theta_2} \right)$$

are derived. These components are substituted into (11), yielding the system

$$\begin{aligned} \frac{d\theta_1}{dr} &= 10(\bar{x}_s - x_s)s_1 + (\bar{\theta}_1 - \theta_1) \\ \frac{d\theta_2}{dr} &= 10(\bar{x}_s - x_s)s_2 + (\bar{\theta}_2 - \theta_2) \\ \frac{dx_s}{dr} &= \sum_{i=1}^2 10(\bar{x}_s - x_s)s_i^2 + \sum_{i=1}^2 (\bar{\theta}_i - \theta_i)s_i \\ &\quad + \lambda(\theta_2 \xi - (\theta_1^2 + \theta_2)x_s), \end{aligned} \quad (17)$$

with initial conditions $\theta_1(0) = \theta_{1,0}$, $\theta_2(0) = \theta_{2,0}$ and $x_s(0) = x_{s,0}$. It can be verified that the objective function J is locally strictly convex in θ – the parameters are locally identifiable – and that the model (14) is asymptotically stable. Accordingly, system (17) converges to a local optimum of the constrained optimisation problem (16) (Theorem 3.1). As stopping criterion for the simulation-based optimisation we use $\max\{\|d\theta/dr\|, \|dx_s/dr\|\} < TOL$ with $TOL = 10^{-6}$. Additionally, a maximal number of function evaluations is implemented.

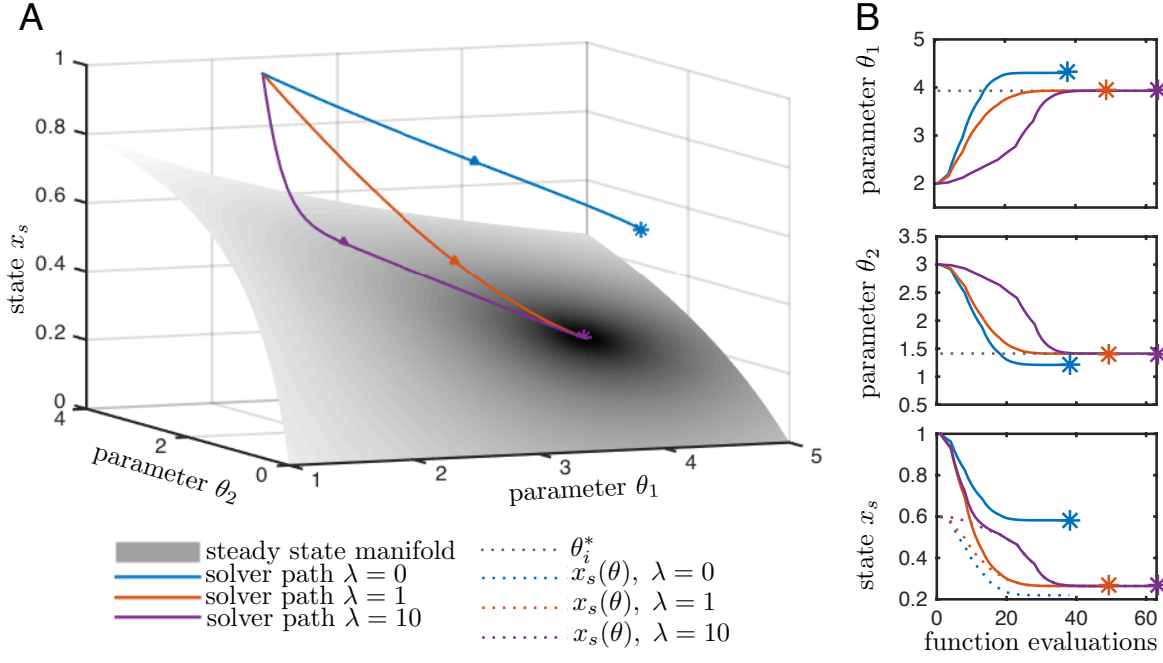


Figure 2: Trajectory of the optimiser path for the estimation of the parameters of the conversion reaction model (14). (A) The trajectory of the system (17) for different retraction factors λ (full lines), the endpoints (stars), and the steady state manifold (surface, (15)) are shown. The objective function values are indicated by the surface coloring. (B) The trajectory of the individual components of system (17) (full lines), their endpoints (stars), and optimal parameter value respective true steady state (15) for the parameters (dotted lines) are depicted.

4.4 Numerical results.

To illustrate the simulation-based optimisation method we simulate system (17) using the MATLAB ODE solver `ode15s` with default settings. Exemplary trajectories are depicted in Figure 2. We find that for retraction factors $\lambda > 0$, the states $(\theta_1, \theta_2, x_s)^T$ converge to the optimal solution. As retraction renders the steady state manifold (15) attractive, also for initial conditions $(\theta_{1,0}, \theta_{2,0}, x_{s,0})^T$ which do not fulfil the steady state condition, fast convergence to the steady state manifold can be achieved using $\lambda \gg 1$. For large retraction ($\lambda \gg 1$), the dynamic consists of two phases: (Phase 1) the state x converges quickly to the parameter-dependent steady state (15); and (Phase 2) the state $(\theta_1, \theta_2, x_s)^T$ moves along the steady state manifold to a local optimum.

5 Example 2: NGF-induced Erk activation

To challenge the proposed simulation-based optimisation method, we consider in the following a larger and more realistic problem: Fitting of dose response data for NGF-induced Erk activation.

5.1 Pathway model.

NGF is known to induce a strong pain sensitisation during inflammation via the activation of Erk. As this is clinically relevant, a simple model for NGF-induced Erk activation has been developed in [10],

$$\begin{aligned} \frac{dx_1}{dt} &= 10^{\theta_1} u (10^{\theta_5} - x_1) - 10^{\theta_2} x_1 \\ \frac{dx_2}{dt} &= (x_1 + 10^{\theta_3}) (10^{\theta_6} - x_2) - 10^{\theta_4} x_2 \end{aligned} \quad (18)$$

in which x_1 denotes the activity of the NGF receptor TrkA, x_2 denotes the activity of Erk, and u denotes the concentration of NGF. The parameters $\theta \in \mathbb{R}^6$ are logarithms of effective rate parameters. The steady state of

model (18) is

$$\begin{aligned} x_{s,1}(\theta) &= \frac{10^{\theta_1 + \theta_5} u}{10^{\theta_1} u + 10^{\theta_2}} \\ x_{s,2}(\theta) &= \frac{10^{\theta_6} (x_{s,1}(\theta) + 10^{\theta_3})}{x_{s,1}(\theta) + 10^{\theta_3} + 10^{\theta_4}} \end{aligned} \quad (19)$$

For details on the model, we refer to the original publication [10].

5.2 Parameter estimation problem.

The parameters θ are unknown and have to be estimated from experimental data. In practice, the activity of the NGF receptor TrkA (x_1) cannot be assessed and merely data for the Erk activity level (x_2) are available. These data can be time resolved, but mostly dose response curves for the steady state are recorded.

In the following, we assume that for ten different NGF concentrations (u_i , $i = 1, \dots, 10$) the stationary Erk activity ($\bar{x}_{s,2}^i$, $i = 1, \dots, 10$) has been measured. Therefore, we evaluate the steady state (19) for these different inputs u_i . Using a standard least squares objective function, we obtain the constrained optimisation problem

$$\begin{aligned} \min_{\theta, \{x_s^i\}_i} J(\theta, \{x_s^i\}_i) &= \frac{1}{2} \sum_{i=1}^{10} (x_{s,2}^i - \bar{x}_{s,2}^i)^2 \\ \text{s.t. } 0 &= 10^{\theta_1} u_i (10^{\theta_5} - x_{s,1}^i) - 10^{\theta_2} x_{s,1}^i \\ 0 &= (x_{s,1}^i + 10^{\theta_3}) (10^{\theta_6} - x_{s,2}^i) - 10^{\theta_4} x_{s,2}^i \\ &i = 1, \dots, 10. \end{aligned} \quad (20)$$

This constrained optimisation problem has $2 \cdot 10 + 6 = 26$ optimisation variables. As an analytical expression for the steady state (19) is available the constrained optimisation problem (20) can be reduced to an unconstrained optimisation problem by substituting the expression for the parameter dependent steady state (20) into the objective function $J(\theta, x_s)$, yielding

$$\min_{\theta} \frac{1}{2} \sum_{i=1}^{10} \left(\frac{10^{\theta_6} \left(\frac{10^{\theta_1 + \theta_5} u_i}{10^{\theta_1} u_i + 10^{\theta_2}} + 10^{\theta_3} \right)}{\frac{10^{\theta_1 + \theta_5} u_i}{10^{\theta_1} u_i + 10^{\theta_2}} + 10^{\theta_3} + 10^{\theta_4}} - \bar{x}_{s,2}^i \right)^2. \quad (21)$$

This unconstrained optimisation problem has 6 optimisation variables, and the size is therefore independent of the number of NGF concentrations.

5.3 Formulation of system for simulation-based optimisation.

As before, a dynamical system converging to the locally optimal points of the constrained optimisation problem (20) can be formulated. To this end the objective function gradient, $dJ/d\theta$, and the local sensitivities, $S(x, \theta)$, are derived. This can be done manually, but for high-dimensional problems symbolic math toolboxes should be used. The results are inserted in (11) to yield a problem specific system.

5.4 Numerical results.

In the following, we compare our simulation-based optimisation method with two state-of-the-art implementations of standard methods:

- Constrained optimisation (20) using the MATLAB algorithm `fmincon` for constrained nonlinear optimisation; and
- Unconstrained optimisation (21) using the MATLAB algorithm `fminunc` for unconstrained nonlinear optimisation.

Constrained optimisation using `fmincon` (or a similar algorithm) is the naïve approach. This approach does neither require an analytical formula for the steady state nor does it exploit the structure of the problem. Unconstrained

optimisation (21) on the other hand requires an analytical solution of the steady state. Using this analytical solution the problem is simplified significantly. Accordingly, we expect that unconstrained optimisation will provide a very good performance, its applicability is however limited to simple models for which an analytical expression for the steady state is known. For the evaluation, `fmincon` is supplied with the objective function value and the values of the constraint, as well as the respective analytical derivatives. For `fminunc` only the objective function and its gradient are implemented. Both algorithms use default settings. The ODE system obtained for the simulation-based method is simulated using the MATLAB solver `ode15s` with default settings. Artificial data for the comparison are generated by calculating the analytical steady state (19) for inputs $u = (0, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100)$ and adding normally distributed noise with zero mean and variance of 0.01.

We used constrained optimisation, unconstrained optimisation as well as the proposed simulation-based method to estimate the parameters. For all methods the initial parameter estimates were sampled from a uniform distribution, $\theta_k \sim \mathcal{U}(-3, 1)$. In addition to initial parameters, constrained optimisation and simulation-based methods require initial estimates for the steady states. They are also sampled from a uniform distribution, $x_{s,j}^i \sim \mathcal{U}(0, 3)$ for $j = 1, 2$ and $i = 1, \dots, 10$.

To assess the convergence properties and the computation time, all aforementioned methods were initialised with 100 sampled starting point. The final objective function values as well as the respective computation time is depicted in Figure 3. We find – as expected – that the unconstrained optimisation method is on average computationally most efficient and for almost all runs a same final objective function value is achieved (which is probably the global optimum). The simulation-based optimisation methods with retraction factor $\lambda = 2$ and $\lambda = 20$ possess similar convergences properties (Figure 3A), the computation time is however roughly 10-times higher. The standard constrained optimisation method implemented by `fmincon` possesses a significantly worse convergence rate of roughly 28%. The computation time for `fmincon` varies over several orders of magnitude and the average is comparable to the computation time for the simulation-based optimisation methods (Figure 3B). However, the mean computation time per converged start is roughly 7 to 8 times higher for `fmincon` in comparison to the simulation based method with retraction factor $\lambda = 20$ (`fmincon`: 7.56 seconds per converged start; simulation based with $\lambda = 20$: 0.98 seconds per converged start).

In general it is difficult to ensure that the objective function is locally strictly convex – a prerequisite for Theorem 3.1 – prior to the optimisation. Also for this example, this was not possible, and indeed it is also not the case. The optimisation results reveal that the parameters are practically non-identifiable and that there is a continuum of parameters which yields the same optimal objective function value. As the simulation-based methods worked fine despite the lack of identifiability, it is well suited for practical applications and we expect that Theorem 3.1 can be extended.

In summary, for this estimation problem, the simulation-based method outperforms standard constrained optimisation methods and achieves a performance only slightly worse than unconstrained optimisation methods. Furthermore, simulations-based methods seem to be robust with respect to non-identifiability. While we use artificial data to evaluate our method, similar experiment data are available (see [10]). This renders also this application example important for practice.

6 Discussion and Outlook

Optimisation problems with steady state constraints are often challenging. If an analytical expression for the parameter and input dependent steady state, $x_s(\theta, u)$ is not available, the vector of optimisation variables contains the unknown parameters as well as the corresponding steady states. Accordingly, optimisers have to evolve on the non-linear manifold defining the set of steady states. In this manuscript, we propose a continuous analogue of a gradient descent algorithm for manifold optimisation. This simulation-based method exploits the local geometry of the steady state manifold for optimisation. The local asymptotic stability of the steady state is used to render the steady state manifold attractive. We establish a result for local convergence using perturbation theory.

The proposed simulation-based optimisation method is evaluated using two models for biological processes. We find that the simulation-based optimisation method possesses improved convergence properties in comparison to standard constrained optimisation methods implemented in the MATLAB routine `fmincon`. Furthermore, the simulation-based optimisation method yields convergence properties almost comparable to those of unconstrained optimisation methods exploiting an analytical expression for the steady state. The proposed simulation-based optimisation method is however applicable to a broader class of problems as an analytical expression for the steady

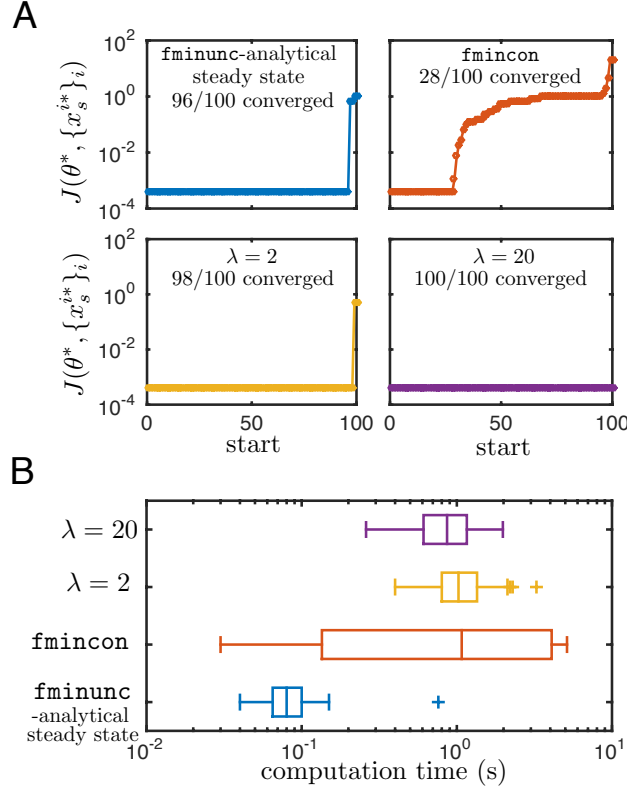


Figure 3: Comparison optimisation methods for NGF-induced Erk activation model. (A) Final objective function values and (B) computation time for 100 runs of unconstrained optimisation method (**fminunc**), constrained optimisation algorithm (**fmincon**) and the proposed simulation-based optimisation algorithm with $\lambda = 2$ and $\lambda = 20$ are depicted.

state is not necessary.

An open question is how the proposed method behaves in the presence of practical and structural non-identifiability. Preliminary results suggest that the method always provides a point on the non-identifiable subspace but this has to be studied in more detail. As in traditional optimisation, also using alternative update schemes, e.g. Newton-type instead of gradient descent-based parameter updates, might yield even more convincing convergence results. As shown in the proof, the retraction factor λ has to be chosen large to ensure convergence. However, a large λ will render the simulation-based system stiff. An intelligent choice of λ is therefore necessary. Furthermore, extension towards a combination of steady state and time series data might be interesting. The simulations-based approach can in principle be generalised to other model classes, including partial differential equations.

References

- [1] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.
- [2] A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller, and J. Timmer, “Lessons learned from quantitative dynamical modeling in systems biology,” *PLoS ONE*, vol. 8, no. 9, e74335, Sept. 2013.
- [3] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. 41 William Street, Princeton, New Jersey: Princeton University Press, 2007.
- [4] D. P. Bertsekas, *Nonlinear programming*, 2nd ed. Athena Scientific, 1999.
- [5] H.-B. Dürr and C. Ebenbauer, “A smooth vector field for saddle point problems,” in *Proceedings of the 50th Conference on Decision and Control (CDC 2011)*, Orlando, Florida, USA, Dec. 2011, pp. 4654–4660.

- [6] K. Tanabe, “Continuous Newton-Raphson method for solving an underdetermined system of nonlinear equations,” *Nonlinear Analysis: Theory, Methods & Applications*, vol. 3, no. 4, pp. 495–503, 1979.
- [7] —, “Global analysis of continuous analogues of the Levenberg-Marquardt and Newton-Raphson methods for solving nonlinear equations,” *Ann. Inst. Statist. Math.*, vol. 37, no. Part B, pp. 189–203, 1985.
- [8] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.
- [9] S. Hug, A. Raue, J. Hasenauer, J. Bachmann, U. Klingmüller, J. Timmer, and F. J. Theis, “High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling,” *Mathematical Biosciences*, vol. 246, no. 2, pp. 293–304, Nov. 2013.
- [10] J. Hasenauer, C. Hasenauer, T. Hucho, and F. J. Theis, “ODE constrained mixture modelling: A method for unraveling subpopulation structures and dynamics,” *PLoS Comput. Biol.*, vol. 10, no. 7, p. e1003686, July 2014.